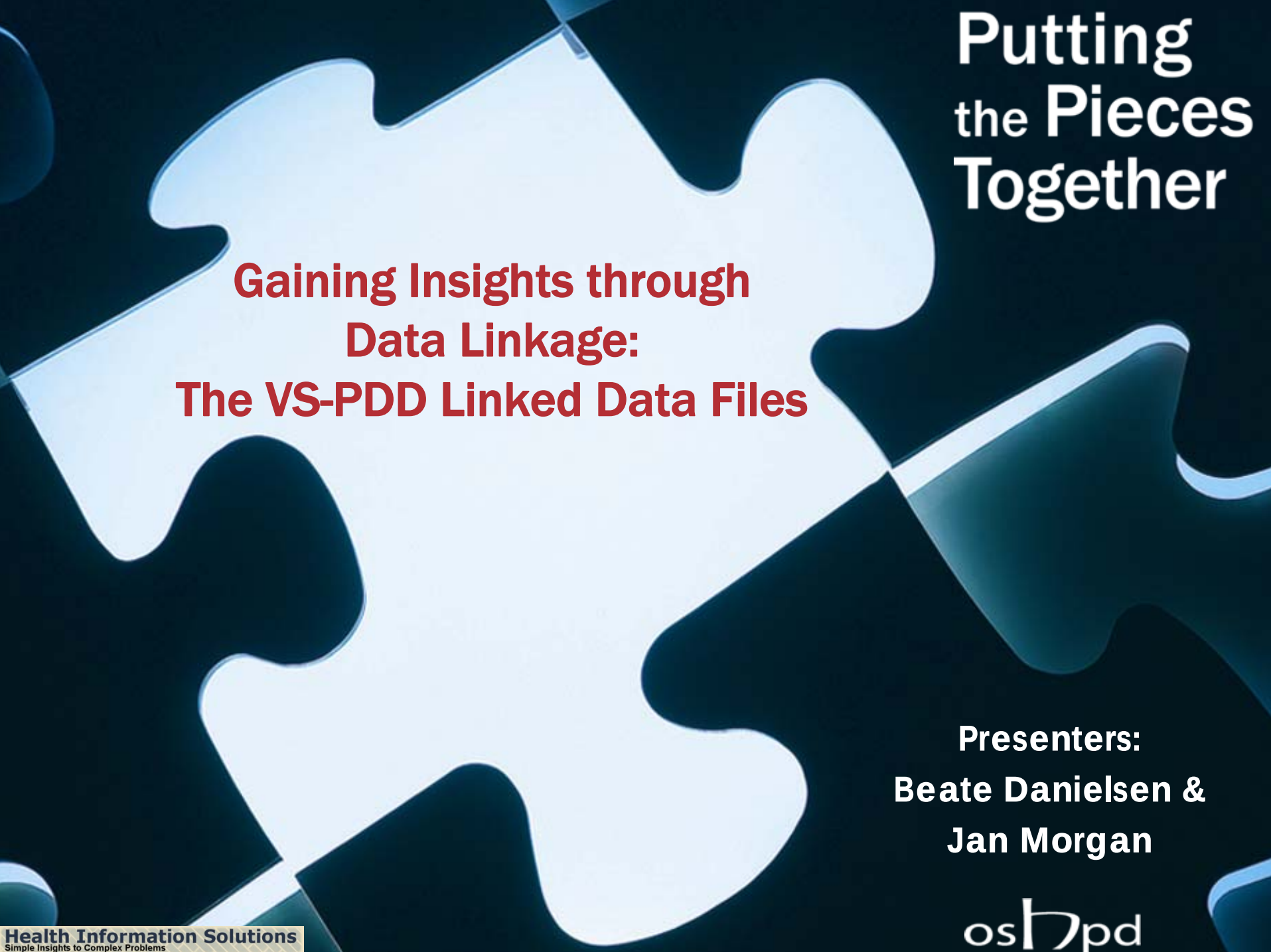

Gaining Insights through Data Linkage: The VS-PDD Linked Data Files

Presenters:
Beate Danielsen &
Jan Morgan

oshpd

Health Information Solutions
Simple Insights to Complex Problems

Goal

Linkage of:

- Vital Statistics Birth Data
- Vital Statistics Fetal Death Data
- Vital Statistics Death Data
- OSHPD Newborn Discharge Data
- OSHPD Maternal Delivery Data
- OSHPD Infant Encounters within First Year (Inpatient, Ambulatory Surgery, Emergency Department)
- OSHPD Maternal Prenatal & Postnatal Encounters (Inpatient, Ambulatory Surgery, Emergency Department)

Vital Statistics Birth Cohort File combines all three.

Structure of Presentation

- Why do we want to link these data sets?
- What are problems and how are they resolved?
- What is the result of the linkage? What percentage of records is successfully linked?
- What data are currently available from OSHPD?
- How can the data be obtained from OSHPD?
- What are core variables to include in your OSHPD data request?
- Summary
- Questions

Why Should we link the Vital Statistics and OSHPD Data?

Vital Statistics Data

- Socio-Demographics
- Prenatal Care
- Delivery Mode
- Mortality Outcomes
- Other Birth Outcomes (Birth Weight, Gestational Age, etc.)

OSHPD Data

- Demographics
- Delivery Mode
- Diagnoses
- Health Care Resource Use Outcomes (Length of Stay, Charges)
- Procedures

Problems

- Different Data Sets with Different Purposes
- No Universal Identifier
- Coding Errors
- Duplicates
- Task size

Problem 1: Different Data Set Owners and Purposes

Vital Statistics Birth, Fetal Death, Death Data:	Maintained by California Department of Public Health for the purpose of vital statistics assessment and health outcomes monitoring
Inpatient Discharge, Ambulatory Surgery, Emergency Department Data:	Maintained by the Office of Statewide Health Planning and Development for analyses contributing to informed decisions on healthcare policy and planning.

Which records can be linked?

“Unlinkable” Records

- Births in locations not reporting to OSHPD
  - Births in Military Hospitals
  - Births in Free-Standing Birthing centers
  - Births at home
- Fetal Deaths
  - Cannot be matched to a newborn discharge record as only live births are admitted as a California inpatient
  - Can be matched to a maternal delivery record

Problem 2: No Universal Identifier

Solution:

Use probabilistic linkage techniques that allow the identification of records that are most likely to be matches.

Match Variables for Linkage of VS Births Record and Newborn PDD

Vital Statistics

- Hospital (4-digit code)
- Infant Birth Date
- Infant Sex
- C-Section Delivery (Y/N)
- ZIP Code of Mom's residence
- Payer source for L&D
- Maternal Race/Ethnicity
- Birth Weight

OSHPD

- Hospital (6-digit code)
- Patient Birth Date
- Patient Sex
- C-Section Delivery (Y/N based on ICD-9-CM DX)
- Patient ZIP
- Payer Source for Encounter
- Patient Race/Ethnicity
- Birth Weight (based on ICD-9-CM DX)

Problem 3: Coding Errors

Solution:

Use probabilistic linkage techniques to find the most likely match for a record

Problem 4: Duplicates

Duplicates of concern since eliminating them from the linkage introduces bias

Use randomization strategy

Strategy for Duplicates

4 observations in Vital Statistics Linked Birth/Infant Death file with the SAME value for birth hospital, ZIP, birth date, sex, race, and payer source:

Case	Birth Weight	Died
1	2,693	No
2	1,814	No
3	3,544	No
4	601	Yes

4 observations in Hospital Discharge File with the SAME value for birth hospital, ZIP, birth date, sex, race, and payer source

Case	ICD9-CM Birth Weight	Died	DRG
1	500 to 750	No	385
2	Normal	No	391
3	Normal	No	390
4	1,750 to 1,999	Yes	388

Linkage Algorithm

Case	Birth Weight	Died	Strategy	ICD-9-CM Birth Weight	Died	DRG
1	2,693	No	Randomization	Normal	No	391
2	1,814	No	Birth Weight	1,750 to 1,999	No	388
3	3,544	No	Randomization	Normal	No	390
4	601	Yes	Birth Weight	500 to 750	Yes	385

Goal of the data linkage is to obtain a functional data set that will allow population-based studies of risks and outcomes using demographic, prenatal, etc., control variables.

The linked data sets cannot be used to track individual cases.

Challenge: Task Size

Source File	Number of Records (2006)
Vital Statistics Birth/Death/Fetal Deaths	567,572 *
Newborn Discharge Record	552,467
Maternal Delivery Record	543,946
Prenatal / Postpartum Encounter (Maternal)	
Inpatient	56,984
Ambulatory Surgery	42,036
Emergency Department	321,734
Postnatal Encounter (Infant)	
Inpatient	48,203 **
Ambulatory Surgery	10,384 **
Emergency Department	355,094 **

* Includes "unlinkable" records

** All records for under 1-year olds born in 2006

Linkage Percentages

Source File	N and % Linked (2006)
Vital Statistics Birth/Death/Fetal Deaths	541,608 / 95.4 % *
Newborn Discharge Record	541,608 / 98.0 %
Maternal Delivery Record	541,608 / 98.1 %
Prenatal / Postpartum Encounter (Maternal)	
Inpatient	Cannot be obtained
Ambulatory Surgery	Cannot be obtained
Emergency Department	Cannot be obtained
Postnatal Encounter (Infant)	
Inpatient	47,031 / 97.6 % **
Ambulatory Surgery	9,560 / 92.1 % **
Emergency Department	344,567 / 97.0 % **

* Includes “unlinkable” records

** Relative to all records for under 1-year olds born in 2006

What Data are Currently Available?

- Linked data for 1991 to 2006
  - 2005 and 2006 linked data include ambulatory surgery and emergency department encounters
  - 2006 data are based on vital statistics birth, vital statistics death, and vital statistics fetal death file since the birth cohort file for 2006 has not yet been published
- Maternal deaths for 2004 to 2006
  - Available as separate files

Data Requests

- Data requests should be directed to the OSHPD – Healthcare Information Division (HID)
- Contact LOUISE HAND – OSHPD/HID/HIRC
- Telephone: (916) 326-3813
- E-mail: LHand@oshpd.ca.gov
- Website: www.oshpd.ca.gov (http://www.oshpd.ca.gov/)
- For web issues contact: OSHPDWebmaster@oshpd.ca.gov

Core Variables Needed to Work with Linked Data

- Except for linked maternal deaths files, linked data are provided as one file per year
- Core variables have been added to these files to ease their use

_brthid

_brthid	_brthidHST	_input	_linkedB	pat_typeI	pat_typeM	_diffI	_diffM	bthwght	diagI00	diagM00
ID of Mom/Baby Pair	ID of Mom/Baby Pair Over Time	Source of Record	Linkage Status of Birth Record	Baby's Encounter Type	Mom's Encounter Type	# Days between birth and encounter	# Days between birth and mom's encounter	Birth Weight	Principal Baby DX	Principal Mom DX
B2001_1	B2001_1	B	Y			0	0	3118	V3000	650
B2001_1	B2001_1	I				308	.	.	486	
B2005_1	B2001_1									650
B2006_1	B2001_1									003
B2006_1	B2001_1									331
B2006_1	B2001_1									
B2006_1	B2001_1									642
B2006_1	B2001_1									003
B2006_1	B2001_1									

- Unique ID assigned to each mom/baby pair for each yearly file.
- Identifies all encounters of mom and baby in discharge, ambulatory surgery (2005 or later), and emergency department (2005 or later) data
- For sets of multiples, each baby has a separate ID

_brthidHST:

_brthid	_brthidHST	_input	_linkedB	pat_typeI	pat_typeM	_diffI	_diffM	bthwght	diagI00	diagM00
ID of Mom/Baby Pair	ID of Mom/Baby Pair Over Time	Source of Record	Linkage Status of Birth Record	Baby Encounter Type	Mom Encounter Type	# Days between birth and encounter	# Days between birth and mom's encounter	Birth Weight	Principal Baby DX	Principal Mom DX
B2001_1	B2001_1	B	Y			0	0	3118	V3000	650
B2001_1	B2001_1	I				308			486	
B2005_1	B2001_1	B								
B2006_1	B2001_1	M								3
B2006_1	B2001_1	B								1
B2006_1	B2001_1	I								
B2006_1	B2001_1	M								2
B2006_1	B2001_1	M								3
B2006_1	B2001_1	I								

- Unique ID assigned to each mom over time.
- Identifies all encounters of mom in discharge, ambulatory surgery (2005 or later), and emergency department (2005 or later) data
- Sets of multiples have the same _brthidHST in common

_input:

_brthid	_brthidHST	_input	_linkedB	pat_typeI	pat_typeM	_diffI	_diffM	bthwght	diagI00	diagM00
ID of Mom/Baby Pair	ID of Mom/Baby Pair Over Time	Source of Record	Linkage Status of Birth Record	Baby Encounter Type	Mom Encounter Type	# Days between birth and encounter	# Days between birth and mom's encounter	Birth Weight	Principal Baby DX	Principal Mom DX
B2001_1	B2001_1	B	Y							50
B2001_1	B2001_1	I								
B2005_1	B2001_1	B	Y							50
B2006_1	B2001_1	M								003
B2006_1	B2001_1	B	Y							331
B2006_1	B2001_1	I								
B2006_1	B2001_1	M								42
B2006_1	B2001_1	M								003
B2006_1	B2001_1	I		E		205	.	.	78703	

Indicates the current type of record
- B: birth/newborn/delivery record
- I: Encounter of infant after birth (transfer, inpatient admission, ED or AS encounter)
- M: Encounter of mom in the prenatal or postpartum period

_linkedB:

_brthid	_brthidHST	_input	_linkedB	pat_t
ID of Mom/Baby Pair	ID of Mom/Baby Pair Over Time	Source of Record	Linkage Status of Birth Record	Ba Enco Ty
B2001_1	B2001_1	B	Y	
B2001_1	B2001_1	I		
B2005_1	B2001_1	B	Y	I
B2006_1	B2001_1	M		
B2006_1	B2001_1	B	Y	I
B2006_1	B2001_1	I		E
B2006_1	B2001_1	M		
B2006_1	B2001_1	M		
B2006_1	B2001_1	I		E

Linkage status for birth/newborn delivery record

Value	VS Birth	Newborn PDD	Maternal PDD
Y	X	X	X
M	X		X
I	X	X	
N		X	X
C			X
B	X		
A		X	

pat_typeI & pat_typeM:

_brthid	_brthidHST	_input	_linkedB	pat_typeI	pat_typeM	_diffI	_diffM	bthwght	diagI00	diagM00
ID of Mom/Baby Pair	ID of Mom/Baby Pair Over Time	Source of Record	Linkage Status of Birth Record	Baby Encounter Type	Mom Encounter Type	# Days between birth and er	# Days between birth and	Birth Weight	Principal Baby DX	Principal Mom DX
B2001_1	B2001_1	B	Y							
B2001_1	B2001_1	I								
B2005_1	B2001_1	B	Y	I	I					
B2006_1	B2001_1	M			E					
B2006_1	B2001_1	B	Y	I	I					
B2006_1	B2001_1	I		E						
B2006_1	B2001_1	M			E					
B2006_1	B2001_1	M			E	.	202	.		64003
B2006_1	B2001_1	I		E		205	.	.	78703	

- Indicate the type of the current OSHPD record
  - I: Inpatient
  - A: Ambulatory Surgery
  - E: Emergency Department
- New variables for 2005 and later

_diffI & _diffM:

_brthid	_brthidHST	_input	_linkedB	pat_typeI	pat_typeM	_diffI	_diffM	bthwght	diagI00	diagM00
ID of Mom/Baby Pair	ID of Mom/Baby Pair Over Time	Source of Record	Linkage Status of Birth Record	Baby Encounter Type	Mom Encounter Type	# Days between birth and encounter	# Days between birth and mom's encounter	Birth Weight	Principal Baby DX	Principal Mom DX
						0	0	3118	V3000	650
						308	.	.	486	
					I	0	0	3033	V3000	650
					E	.	-180	.		64003
					I	0	-1	2807	V3000	66331
						115	.	.	3829	
					E	.	200	.		V642
					E		202	.		64003
						205	.	.	78703	

- Number of days between baby (_diffI) or mom (_diffM) encounter (admission date) and birth
- Negative numbers correspond to prenatal encounters
- Positive numbers correspond to postnatal encounters

bthwght, diagI00, & diagM00:

_brthid	_brthidHST	_input	_linkedB	pat_typeI	pat_typeM	_diffI	_diffM	bthwght	diagI00	diagM00
ID of Mom/Baby Pair	ID of Mom/Baby Pair Over Time	Source of Record	Linkage Status of Birth Record	Baby Encounter Type	Mom Encounter Type	# Days between birth and encounter	# Days between birth and mom's encounter	Birth Weight	Principal Baby DX	Principal Mom DX
B2001_1	B200						0	3118	V3000	650
B2001_1	B200						.	.	486	
B2005_1	B200						0	3033	V3000	650
B2006_1	B200						-180	.		64003
B2006_1	B200						-1	2807	V3000	66331
B2006_1	B200						.	.	3829	
B2006_1	B200						200	.		V642
B2006_1	B200						202	.		64003
B2006_1	B200						.	.	78703	

- Example of linked information:
  - bthwght: Birth weight from vital statistics data
  - diagI00: Principal DX for baby encounter
  - diagM00: Principal DX for mom encounter
- Information from all three sources only present for linked birth records

_twinwght:

_brthid	_linkedI	_input	_twinwght	pat_typeI	pat_typeM	_diffI	_diffM	bthwght	typebth	diagI00	diagM00
ID of Mom/Baby Pair	Baby Encounter ID	Source of Record	Multiples Weight	Baby Encounte Type			# Days				al X
B2006_2		M	0								
B2006_2		M	0								
B2006_2		M	0								
B2006_2	2006_1	B	0	I							
B2006_2	ED06_1	I	.	E							
B2006_2	ED06_2	I	.	E							
B2006_2	ED06_3	I	.	E							
B2006_3		M	1								
B2006_3		M	1								
B2006_3		M	1								
B2006_3	2006_2	B	1	I							
B2006_3	ED06_4	I	.	E							
B2006_3	ED06_5	I	.	E							
B2006_3	ED06_6	I	.	E							

- The variable _twinwght is 1 for one infant in a set of multiples; for all other infants in the same set of multiples, _twinwght is 0.
- Identify sets of multiples delivered by the same mother
- Generate a correct count of deliveries. For instance, in order to obtain the average maternal age including multiple births all _input EQ 'B' records should be used using _twinwght as weight for each observation in the data set.

_twinwght:

_brthid	_linkedI	_input	_twinwght	pat_typeI	pat_typeM	_diffI	_diffM	bthwght	typebth	diagI00	diagM00
ID of Mom/Baby Pair	Baby Encounter ID	Source of Record	Multiples Weight	Baby Encounter Type	Mom Encounter Type	# Days between birth and encounter	# Days between birth and mom's encounter	Birth Weight	Type of Birth	Principal Baby DX	Principal Mom DX
B2006_2		M	0		E	.	-361	.			6825
B2006_2		M	0		E	.	-318	.			78650
B2006_2		M	0		E	.	-222	.			64893
B2006_2	2006_1	B	0	I	I	0	0	2523	2	V3101	65101
B2006_2	ED06_1	I	.	E		31	.	.		V719	
B2006_2	ED06_2	I	.	E		112	.	.		56400	
B2006_2	ED06_3	I	.	E		318	.	.		78703	
B2006_3		M	1		E	.	-361	.			6825
B2006_3		M	1		E	.	-318	.			78650
B2006_3		M	1		E	.	-222	.			64893
B2006_3	2006_2	B	1	I	I	0	0	2608	2	V3101	65101
B2006_3	ED06_4	I	.	E		31	.	.		V719	
B2006_3	ED06_5	I	.	E		210	.	.		7806	
B2006_3	ED06_6	I	.	E		234	.	.		7849	

Summary

- Linkage task successfully accomplished using probabilistic match techniques
- No evidence of bias introduced by the linkage process
- Usage of randomization minimally affects population-based statistics
- Algorithm is regularly updated to account for changes in the structure of the input data or improved efficiency
- The resulting data set is suitable for population-based studies
- Linkage results available for download at http://www.health-info-solutions.com

Questions?